AUSTRALASIAN JOURNAL OF COMBINATORICS
Volume **78(2)** (2020), Pages 329–334

Short witnesses for Parikh-friendly permutations

Jamie Simpson

Department of Mathematics and Statistics
Curtin University of Technology
GPO Box U1987, Perth, WA 6845
Australia
Jamie.Simpson@curtin.edu.au

Abstract

Recently Wen Chean Teh [*Australas. J. Combin.* 76 (2020), 208–219] showed that for every permutation π of the ordered alphabet $A = \{a_1, a_2, \ldots, a_n\}$ there exists a word $w \in A^*$, in which each letter of A appears at least once, such that w and $\pi(w)$ have the same Parikh matrix. He conjectured that it is always possible to find such a w whose length is at most $2n$. We prove this.

We use the usual notation for combinatorics on words. A word of m elements is $x = x[1\,..\,m]$, with $x[i]$ being the ith element and $x[i\,..\,j]$ the *factor* of elements from position i to position j. If $i = 1$ then the factor is a *prefix* and if $j = m$ then it is a *suffix*. The *reverse* of x, written $\overline{x}$, is the word $x[m]x[m-1]\ldots x[1]$. If a word equals its own reverse then it is a *palindrome*. The letters in x come from some *alphabet*. We will use the *ordered* alphabet $A = \{a_1, a_2, \ldots, a_n\}$ where $a_1 < a_2 < \cdots < a_n$, except in some examples where we use the more familiar $a < b < c < \cdots$. The set of all finite words with letters from A is A^*. The *length* of x, written $|x|$, is the number of letters in x. If $i \leq j$ we use $a_{i,j}$ as an abbreviation for $a_i a_{i+1} \ldots a_j$ with $a_{i,i}$ interpreted as a_i. If $u_1, u_2, \ldots, u_k, v_1, v_2, \ldots, v_k$ are words and $w = u_1 v_1 u_2 v_2 \ldots u_k v_k$ then $u_1 u_2 \ldots u_k$ is a *subword* (sometimes called a *scattered subword*) of w. The number of occurrences of u as a subword of w is written $|w|_u$. For example the word $w = cbabbacb$ contains 4 occurrences of the subword ab so $|w|_{ab} = 4$. The number of occurrences of the single letter c in w is $|w|_c = 2$. Note that ca is a subword of w but not a factor. The set of letters occurring in a word w is written $alph(w)$.

This note concerns *Parikh Matrices* which may be defined as follows (for an alternative but equivalent definition see [3]). The Parikh matrix for a word w defined on the alphabet $\{a_1, a_2, \ldots, a_n\}$ is an $(n+1) \times (n+1)$ matrix $\Psi(w)$ with entries $\psi_{i,j}$ where

$$\begin{aligned} \psi_{i,j} &= 0 && \text{for } 1 \leq j < i \leq n+1, \\ \psi_{i,i} &= 1 && \text{for } 1 \leq i \leq n+1, \\ \psi_{i,j+1} &= |w|_{a_{i,j}} && \text{for } 1 \leq i \leq j \leq n. \end{aligned}$$

ISSN: 2202-3518 ©The author(s). Released under the CC BY 4.0 International License

Thus

$$\Psi(cbabbacb) = \begin{pmatrix} 1 & 2 & 4 & 2 \\ 0 & 1 & 4 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We will refer to the diagonal immediately above the main diagonal as the second diagonal, the one above that as the third diagonal and so on. The entries in the second diagonal form the Parikh vector of w, $(|w|_{a_1}, |w|_{a_2}, \ldots, |w|_{a_n})$. This gives some information about w, the Parikh matrix gives rather more. Much of the research in the area concerns the conditions under which different words have the same Parikh matrix. In 2004 Şerbănuţă [2] extended the idea of Parikh matrices to involve a permutation π of the alphabet. In this case the Parikh matrix is written $\Psi_{\pi(a_{1,n})}(w)$ and the third condition of the definition is replaced by

$$\psi_{i,j+1} = |w|_{\pi(a_{i,j})}.$$

A further extension is possible in which π is replaced be a word in A^* but this will not concern us. The non-extended version of the Parikh Matrix corresponds to π being the identity permutation.

In 2018 Salomaa [1] defined a permutation π to be Parikh friendly with respect to A if there exists a word w with $alph(w) = A$ and such that

$$\Psi_{a_1..a_n}(w) = \Psi_{\pi(a_1..a_n)}(w)$$

and the word w is then said to be a *witness* for the permutation π. But for any $a_{i,i+k}$ we have $|w|_{\pi(a_{i,i+k})} = |\pi^{-1}(w)|_{a_{i,i+k}}$, so w is a witness of π if and only if

$$\Psi_{a_1..a_s}(w) = \Psi_{a_1..a_s}(\pi^{-1}(w)). \tag{1}$$

We are now dealing with the non-extended version so the subscripts can be dropped. Salomaa asked for a characterisation of Parikh-friendly permutations. In [3] Teh obtained the surprising result that every permutation π of A is Parikh-friendly. In fact he proved the stronger result that for every alphabet A there exists a single (long) word (which he called a *uniformly Parikh-friendly word with respect to* A) that is a witness for every permutation of A. Teh also made the following conjecture which we will prove in Corollary 7.

Conjecture 1. *Suppose A is an ordered alphabet, then every permutation of A has a Parikh-friendly witness of length at most* $2|A|$.

Since we want witnesses for all permutations, and as π runs through all permutations in the symmetric group so does π^{-1}, so in view of (1) we will find w depending on π such that $\Psi(w) = \Psi(\pi(w))$. Our construction of such a witness is fairly simple but we need some machinery before we present it. Recall that τ is a cyclic permutation if it is a cyclic group under composition: such a permutation can be written $(a, \tau(a), \tau^2(a), \ldots, \tau^{k-1}(a))$ where the terms are distinct and $\tau^k(a) = a$. Any permutation of a finite set can be written as the product of cyclic permutations. We say that a word is *special* if it is a palindrome with each letter in the word occurring exactly twice. Note that if w is special and π is a permutation of $alph(w)$ then $\pi(w)$ is also special. We will show in Theorem 6 that any cyclic permutation has a witness that is special and, using Lemma 5, that the concatenation of the witnesses of the cyclic permutations is a witness for their product. The truth of Conjecture 1 then follows. We begin by describing the Parikh matrix of a special word.

Lemma 2. *If w is special then w contains subwords $a_{i,i+k}$ and $a_{j,j+k}$ where $i < j \leq i + k - 1$ if and only if w contains the subword $a_{i,j+k}$.*

Proof. Sufficiency is immediate. We will prove necessity. Recall that w is a palindrome. It is sufficient to show that there is at least one position in w that is used in both subwords. Both subwords contain the alphabetic letters a_{i+k-1} and a_{i+k}. It is clear that for at least one of these letters, both appearances of the letter correspond to the same position in w. From this it is clear that one can construct the required subword using the given subwords as a prefix and an overlapping suffix. □

The condition $j \leq i + k - 1$ cannot be replaced by $j \leq i + k$. Consider $w = cbaeddeabc$ which contains subwords abc and cde, but not $abcde$.

Theorem 3. *If P is the Parikh matrix defined on A of a special word w then*
(a) Each element on the main diagonal equals 1,
(b) The elements in the second and third diagonals equal 0 *or* 2, *and each element $p_{i,i+2}$ in the third diagonal equals* 2 *if and only if both $p_{i,i+1}$ and $p_{i+1,i+2}$ equal* 2.
(c) Each element on the fourth diagonal equals 0 *or* 2.
(d) Each element $p_{i,j}$ of P with $j \geq i + 4$ equals 2 *if both $p_{i,j-1}$ and $p_{i+1,j}$ equal* 2, *and equals* 0 *otherwise.*

Proof. Part (a) is immediate from the definition of Parikh matrices. For (b) note that if w contains the element $a_i \in A$ then, as w is special, it contains exactly two copies of a_i and so $|w|_{a_i} = p_{i,i+1} = 2$. Note that w contains the subword $a_i a_{i+1}$ if and only if it contains both a_i and a_{i+1}, and it will then contain exactly two copies of the subword. Part (b) follows. Part (c) follows from similar reasoning: if a subword $a_{i,i+2}$ occurs in w it occurs exactly twice. Part (d) follows from Lemma 2: if, and only if, $|w|_{a_{i,i+2}}$ and $|w|_{a_{i+1,i+3}}$ both equal 2 then so does $|w|_{a_{i,i+3}}$, and similarly for longer subwords. □

The essential point of this theorem is the following corollary.

Corollary 4. *In order to show that the Parikh matrices of two special words w_1 and w_2 are equal it is sufficient to show that $alph(w_1) = alph(w_2)$ and that their fourth diagonals are equal.*

Proof. It is clear that the described conditions are necessary. The condition $alph(w_1) = alph(w_2)$ means the second diagonals of the two Parikh matrices are equal and part (b) of the theorem then implies that the third diagonals are equal. Parts (c) and (d) imply that if the fourth diagonals are equal then inductively all higher diagonals are equal and so the matrices are equal. □

Recall that the fourth diagonal contains terms $|w|_{a_{i,i+2}}$ which count the subwords $a_i a_{i+1} a_{i+2}$. We will call such a subword a *consecutive triple.* Note that a consecutive triple is uniquely determined by its first element.

Lemma 5. *If u_1, u_2, v_1 and v_2 are words such that*

$$\Psi(u_1) = \Psi(u_2) \text{ and } \Psi(v_1) = \Psi(v_2)$$

and $alph(u_1) \cap alph(v_1) = \emptyset$ then

$$\Psi(u_1 v_1) = \Psi(u_2 v_2).$$

Proof. Consider $|u_1 v_1|_{a_{i,i+k}}$ and suppose that this is positive. If $a_{i,i+k}$ is a subword of u_1 then, by our assumption that $\Psi(u_1) = \Psi(u_2)$, $a_{i,i+k}$ is a subword of u_2 and $|u_1|_{a_{i,i+k}} = |u_2|_{a_{i,i+k}}$. Since the alphabets of u_1 and v_1 (which equal, respectively, the alphabets of u_2 and v_2) do not intersect this means $|u_1 v_1|_{a_{i,i+k}} = |u_2 v_2|_{a_{i,i+k}}$. For similar reasons this will also be the case if $a_{i,i+k}$ is a subword of v_1.

Suppose instead that $a_{i,i+s}$ is a subword of u_1 and $a_{i+s+1,i+k}$ is a subword of v_1, where $0 \leq s < k$. Since $|u_1|_{a_{i,i+s}}$ must equal $|u_2|_{a_{i,i+s}}$ and $|v_1|_{a_{i+s+1,i+k}}$ must equal $|v_2|_{a_{i+s+1,i+k}}$ we have

$$\begin{aligned} |u_1 v_1|_{a_{i,i+k}} &= |u_1|_{a_{i,i+s}} |v_1|_{a_{i+s+1,i+k}} \\ &= |u_2|_{a_{i,i+s}} |v_2|_{a_{i+s+1,i+k}} \\ &= |u_2 v_2|_{a_{i,i+k}} \end{aligned}$$

in all cases. Finally, suppose that $|u_1 v_1|_{a_{i,i+k}} = 0$. If also $|u_2 v_2|_{a_{i,i+k}} = 0$ there is nothing to prove, and if $|u_2 v_2|_{a_{i,i+k}} > 0$ then we obtain a contradiction by repeating the argument above but beginning with the assumption that $|u_2 v_2|_{a_{i,i+k}}$ is positive. It follows that in this case $\Psi(u_1 v_1) = \Psi(u_2 v_2)$. □

Recall that $\overline{X}$ is the reverse of X.

Theorem 6. *Let $\pi = (b_1, b_2, \ldots, b_t)$ be a cyclic permutation of a subset of A with b_1 being its lexicographically least member and set $X = b_2 b_3 \ldots b_t$. Let $w_1 = b_1 X \overline{X} b_1$ and $w_2 = \overline{X} b_1 b_1 X$. Then for either $i = 1$ or $i = 2$ (but not both) we have*

$$\Psi(w_i) = \Psi(\pi(w_i)).$$

Proof. The words w_1 and w_2 are special so we can apply earlier results to them. Note that

$$\pi(b_1 X) = \pi(b_1 b_2 \ldots b_t) = b_2 b_3 \ldots b_t b_1 = X b_1$$

and

$$\pi(\overline{X} b_1) = \pi(b_t b_{t-1} \ldots b_1) = b_1 b_t \ldots b_2 = b_1 \overline{X}.$$

Consider $w_1 = b_1 X \overline{X} b_1$ and note that $\pi(b_1 X \overline{X} b_1) = X b_1 b_1 \overline{X}$. Clearly $alph(w_1) = alph(w_2)$ so we may apply Corollary 4. If $b_1 X \overline{X} b_1$ and $X b_1 b_1 \overline{X}$ have the same set of consecutive triples then they have the same Parikh matrices and we are done. Clearly any consecutive triple in $X\overline{X}$ will belong to both $b_1X\overline{X}b_1$ and $Xb_1b_1\overline{X}$. Therefore, if there is a consecutive triple in one of these words which is not in the other it must contain b_1, and since we have assumed b_1 is the lexicographically least letter in the permutation the consecutive triple must begin with b_1. Say it is $b_1\alpha\beta$. If either α or β is absent from the set $\{b_1, b_2, \dots, b_t\}$ then $b_1\alpha\beta$ does not appear in either word and we're done. Suppose they are both present. Clearly $b_1\alpha\beta$ is then a subword of $b_1X\overline{X}b_1$. It is also a subword of $Xb_1b_1\overline{X}$ if α occurs before β in $\overline{X}$; that is, if β occurs before α in X. So $\Psi(w_1) = \Psi(\pi(w_1))$ except in the case where $\alpha\beta$ is a subword of X.

Suppose then that $\alpha\beta$ is a subword of X and consider $w_2 = \overline{X}b_1b_1X$. We have $\pi(\overline{X}b_1b_1X) = b_1\overline{X}Xb_1$. As before, any consecutive triple which is a subword of $\overline{X}X$ will be a subword of both $\overline{X}b_1b_1X$ and $b_1\overline{X}Xb_1$. So we need only consider the consecutive triple $b_1\alpha\beta$. By our assumption that $\alpha\beta$ is a subword of X this is indeed a subword of both words, so $\Psi(w_2) = \Psi(\pi(w_2))$ □

If we had written the cyclic permutation as $(c_1, \dots, c_k)$ with c_k being the lexicographically greatest letter, and set $Y = c_1 \dots c_{k-1}$ we could have used the words $w_3 = c_kY\overline{Y}c_k$ and $w_4 = \overline{Y}c_kc_kY$ instead of w_1 and w_2 in the theorem. I haven't found other satisfactory words. We can now prove Conjecture 1.

Corollary 7. *If π is a permutation of $A = \{a_1, \dots, a_n\}$ then π has a witness of length $2n$.*

Proof. The permutation π can be written as a product of disjoint cyclic permutations $C_1, C_2, \dots C_k$ which have lengths $\mu_1, \dots, \mu_k$ respectively, so that $\mu_1 + \mu_2 + \dots + \mu_k = n$. By the theorem for each of these there exists a word w_i of length $2\mu_i$ such that

$$\Psi(w_i) = \Psi(\pi(w_i)).$$

Since the C_is are disjoint the sets $alph(w_i)$ are disjoint so we may apply Lemma 5 and conclude that

$$\begin{aligned} \Psi(w_1 \dots w_k) &= \Psi(\pi(w_1) \dots \pi(w_k)) \\ &= \Psi(\pi(w_1 \dots w_k)). \end{aligned}$$

So the word $w_1 \dots w_k$ is a witness for π and has length $2(\mu_1 + \dots + \mu_k) = 2n$. □

We finish with an example. Consider the permutation $\pi(abcdefgh) = hgafdecb$ which is a product of the cyclic permutations $(ahbgc)$ and (dfe). For the first cyclic permutation, since bc is a subword of $X = hbgc$, we use the word w_2 from Theorem 6 which is $cgbhaahbgc$. For the second permutation ef is not a subword of $X = fe$ so we use $w_1 = dfeefd$. Our witness is the concatenation of these which is $cgbhaahbgcdfeefd$. The image of this under π is $acgbhhbgcafeddef$. The Parikh matrix of each of these is:

$$\begin{pmatrix} 1 & 2 & 2 & 2 & 4 & 4 & 0 & 0 & 0 \\ 0 & 1 & 2 & 2 & 4 & 4 & 4 & 0 & 0 \\ 0 & 0 & 1 & 2 & 4 & 4 & 4 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 2 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Acknowledgements

Thanks to the referees for their helpful suggestions.

References

[1] A. Salomaa, "Parikh Matrices: Subword Indicators and Degrees of Ambiguity", in *Adventures Between Lower Bounds and Higher Altitudes*, H. J. Bockenhauer, D. Komm and W. Unger, Springer, 2018, 100–112.

[2] T. F. Şerbănuţă, Extending Parikh matrices, *Theoretical Computer Science* 310 (2004), 233–246.

[3] W. C. Teh, Parikh-friendly permutations and uniformly Parikh-friendly words, *Australas. J. Combin.* 76 (2020), 208–219.

(Received 16 Jan 2020; revised 6 July 2020)